

**Findind spot shape in cdna microarray by
using a deformable grid and a Markov
segmentation**

C. Gouinaud¹ D. Hill² P. Peyret³ L. Yon⁴

Research Report LIMOS/RR-09-02

30 avril 2009

¹LIMOS-ISIMA, Campus des Cézeaux, 63177 Aubière, gouinaud@isima.fr

²LIMOS-ISIMA, Campus des Cézeaux, 63177 Aubière, drch@isima.fr

³Laboratoire Microorganismes : Genome et Environnement UMR CNRS
6023, Clermont Université, Campus des Cezeaux, 63177 Aubiere Cedex, France.
pipeyret@iut.u-clermont1.fr

⁴LIMOS-ISIMA, Campus des Cézeaux, 63177 Aubière, loic@isima.fr

1 Introduction

L'intérêt de l'utilisation des biopuces cdna pour la génétique n'est plus à démontrer [EISE-99]. Cette technologie complexe arrive maintenant à une certaine maturité et son utilisation s'étend notamment dans la modélisation des relations gène expression individu. De ce fait, le défi actuel est l'amélioration de la précision des mesures réalisées de façon à augmenter la qualité des expressions estimées et donc les résultats fonctionnels. En effet, le plus souvent les réponses cherchées jusqu'à présent étaient binaires, alors que maintenant la recherche s'oriente vers des mesures moins tranchées où l'on veut mesurer un quantum d'expression.

D'un point de vue concret, l'utilisation des biopuces fait appel à des technologies très différentes. Les sources d'imprécision dans le process de production sont nombreuses et il est donc fondamental de maîtriser chaque étape et d'appliquer avec rigueur chaque procédure.

Parmi toutes les étapes, celle du traitement d'images est une des plus importantes et certainement celle où les gains de précision les plus importants peuvent être obtenus [BROW-01] [KATZ-02]. Cette phase est aussi celle où les risques d'erreurs sont les plus grands, car la qualité des mesures réalisées est évidemment affectée par toutes les étapes précédentes. Le travail présenté ici s'attaque donc à l'amélioration du calcul des expressions en proposant une nouvelle méthode de segmentation des images.

La première motivation de ce travail est de construire une méthode de traitement d'images plus générique que celle proposée par les concepteurs de robot spotteur et de scanner à biopuces. En effet, ces derniers sont souvent associés à des logiciels de traitement d'images développés dans le but d'utiliser les métadonnées décrivant la lame fournie par les robots d'un même constructeur [YANG-02]. Ceci complique en particulier les comparatifs entre des biopuces produites avec des appareils différents et donc les échanges de données images entre laboratoires. L'utilisation conjointe de lames produites avec du matériel hétérogène est très complexe et la gestion de l'obsolescence des robots spotteur est un gros problème.

Ce travail vise aussi à arriver à une qualité plus importante de segmentation des images et en particulier à gérer correctement les déformations de spots, à corriger au mieux les déplacements de sonde et donc à fournir les meilleurs masques possibles pour le calcul des expressions malgré les différents bruits présents.

La méthode proposée ici est originale car elle combine des approches différentes pour chaque étape nécessaire à la détection des spots. Elle s'inspire de méthodes utilisées dans d'autres contextes tels que la télédétection sur image SAR qui sont également acquises en lumière cohérente et présentent donc le même genre de bruit [GOOD-76]. Elle présente aussi l'intérêt de faire intervenir au minimum un opérateur tout en tolérant l'utilisation d'images issues de matériel quelconque.

Ce document est présenté en trois parties : la première consacrée au contexte de cette étude, la seconde traitant du problème de positionnement de la grille sur l'image et la dernière présentant notre méthode de segmentation et ses résultats.

2 Contexte

comme cela est dit plus haut, le calcul de l'expression des spots nécessite deux étapes préliminaires de traitement, une de localisation grossière des zones où se trouvent les sondes sur l'image, nous l'appellerons positionnement de la grille, et une réalisant le calcul de la liste des pixels appartenant à un spot, que nous appellerons segmentation. Le positionnement est essentiel pour éviter des erreurs dramatiques d'association risquant de conduire à des résultats erronés d'expérience [WIER-02]. Il est couramment réalisé suivant deux méthodes qui sont l'utilisation de métadonnées ou l'association zone-spot après segmentation de l'image.

- l'utilisation de métadonnées nécessite l'emploi d'appareils et de logiciels d'un unique constructeur ou au minimum une interopérabilité entre tous les acteurs de la chaîne. Le repérage des spots se fait donc a priori en utilisant la position des dépôts fournis par le spoteur et la position de la zone imagée fournie par le logiciel pilotant le scanner. Pratiquement, ces données sont inscrites dans des fichiers qui sont échangés entre les postes servant au traitement. Pour que le positionnement réussisse, il faut que tous les fichiers soient décodés sans erreur par tous les éléments de cette suite de logiciels. Cela présente le défaut principal de condamner l'utilisation de matériel hétérogène et de ne pas tenir compte des variations de position des spots après le dépôt.
- la stratégie inverse de la précédente méthode consiste à réaliser la segmentation de l'image en spot puis à plaquer dessus la grille des sondes fournies par le spoteur [WEE-03]. Le but de cette méthode est de corriger les éventuels effets de glissement des spots ou autres artefacts introduits durant la production de la lame, la phase d'hybridation et l'acquisition de l'image. Il s'agit ici soit de processus d'association individuelle de spots, soit de méthode associant recherche des blocs de dépôt et association des spots dans les blocs. Après l'essai de diverses variantes de cette classe de méthodes [KATZ-03], nous devons constater qu'il s'agit encore d'un problème ouvert. En effet, ces techniques ne peuvent s'affranchir des risques inhérents au glissement de sonde, aux erreurs de sens d'insertion de la lame, ou aux images ne représentant qu'une vue partielle de la lame. Le risque de non détection complète d'une ligne ou d'une colonne de la biopuce ne peut être négligé et il conduit à une incertitude systématique. Ce dernier cas se produit du fait de la non hybridation d'une série de sondes et donc

son avènement est imprévisible. De plus, la nécessité de la prédétection oblige à traiter l'intégralité de l'image et donc introduit un risque important de fausse détection de spot.

Face à ce constat, nous proposons ici une méthode semi-automatique de positionnement, décrite dans la partie 3, permettant par l'intervention très courte d'un opérateur de corriger les défauts de la deuxième classe de méthode de positionnement sans requérir de métadonnées et donc de matériel ou de logiciel particulier.

Le but de la segmentation est de séparer les pixels d'un spot de ceux du fond. Les méthodes les plus couramment employées fonctionnent soit à l'aide d'algorithmes s'appuyant sur la reconnaissance des formes [HIRA-02] ou bien à l'aide de classifications basées sur la valeur des pixels [YIDO-97].

- dans le cas où on utilise des critères basés sur la forme, on présuppose que les spots visibles sur l'image ont des caractéristiques géométriques conformes aux sondes déposées. Les caractéristiques présupposées sont donc le plus souvent la circularité et la convexité du spot. Les techniques employées utilisent soit des cercles fixes, soit des cercles ou ellipses adaptatifs mais, dans tous les cas, le positionnement et la forme sont des présupposés [PET-04]. Bien évidemment, ces conditions ne sont que rarement réalisées et les résultats produits conduisent soit au rejet de nombreux spots soit à la prise en compte de nombreux pixels de fond dans les expressions calculées.
- l'utilisation de classification est basée sur l'existence d'une différence statistique entre les valeurs des pixels de fond et ceux constituant les spots. Ces méthodes sont le plus souvent basées sur l'analyse de statistiques du premier ordre permettant de discriminer localement les pixels du spot de ceux du fond [WEE-03]. Elles sont parfois associées à des techniques de croissance de région de contours actifs qui réintroduisent l'aspect reconnaissance de forme et à partir d'un présupposé fort de positionnement initial.

Nous proposons ici une méthode hybride des deux approches précédentes basée sur une double prise en compte de la forme et des signaux visant à s'affranchir des petites erreurs de positionnement et intégrant des hypothèses de connexité par arc⁵ des spots. D'autres travaux ont développé des approches similaires telles que celles basées sur les classifications de Man-Whitney ou basées sur des techniques d'association dissociation ([BAR-06]). Notre approche se distingue par l'emploi de critères topologiques de voisinage associé une approche statistique évoluée. La partie 4 expose notre approche basée sur une segmentation markovienne.

⁵cela signifie pratiquement que les spots sont considérés comme étant en un seul morceau.

3 Positionnement de la grille

La première des tâches nécessaire à la segmentation de biopuces est le repositionnement de la grille des spots sur l'image scannée [KULK-00]. Ce sujet a fait l'objet de nombreux travaux, utilisant des grilles déformables, basés sur une prédétection des spots ou basés sur la reconnaissance de blocs de spots [GIUL-07].

Cependant, les résultats de ces travaux se confrontent difficilement à la réalité des laboratoires et ont donc du mal à positionner correctement les grilles dès lors que :

- La lame n'est pas mise dans le sens prévu dans le scanner.
- Beaucoup de spots ne sont pas exprimés aussi bien en rouge qu'en vert.
- La résolution du scanner peut être réglée par l'opérateur.
- La zone scannée ne couvre pas la lame et l'information de positionnement ne passe pas entre le logiciel de scan et le logiciel de traitement d'image.

Le dernier problème évoqué ci-dessus se produit dès qu'un laboratoire dispose de plusieurs scanners de marques différentes ou dès que l'on produit son propre logiciel, ce qui est notre cas. Nous cherchons également à améliorer les techniques de segmentation. Il nous faut donc une information fiable de positionnement de façon à éviter des erreurs grossières. D'autre part, dès que les laboratoires travaillent sur de petites séries ou testent différentes configurations, il est difficile pour eux de maintenir des conditions d'expérimentation drastiques et constantes.

3.1 Positionnement global

La double incertitude induite par les spots non exprimés et le positionnement incertain de la lame rendent insoluble la recherche d'un positionnement automatique. Ceci nous a conduit à développer une méthode de positionnement de grille semi-automatique qui, tout étant la moins contraignante possible pour l'opérateur, élimine le plus possible les défauts du positionnement totalement automatique. Pour cela, nous nous sommes inspirés de méthodes utilisées en cartographie. Le positionnement global de la grille se fait donc en désignant des pixels de l'image et leur positions idéales puis en calculant une interpolation relativement à un modèle de déformation.

Le choix du modèle conditionne le nombre de points amers minimum permettant la détermination des paramètres qu'il contient. Ce choix doit être fait en fonction de la physique du système d'acquisition de façon à compenser les déformations par rapport à la prise de vue idéale. Dans le domaine des scanners à objectif confocal pour biopuces, la prise de vue est correcte quand la lame est dans un plan parallèle au déplacement de l'objectif et quand l'orientation de la lame est telle que son bord le plus long est parallèle à la trajectoire du centre de rotation du bras du scanner. Dans ce cas, les lignes et les colonnes de l'image obtenue

coïncident avec les axes de la grille issus du spoteur. Les déformations globales de l'image sont donc de deux natures possibles suivant que l'on considère que le plan de la lame peut changer ou pas. Dans ce dernier cas, les déformations de l'image seront modélisables sous forme d'une combinaison d'applications affines en 2D dans le plan de l'image alors que dans le premier cas, il s'agira de modéliser un déplacement 3D dont on observe une projection.

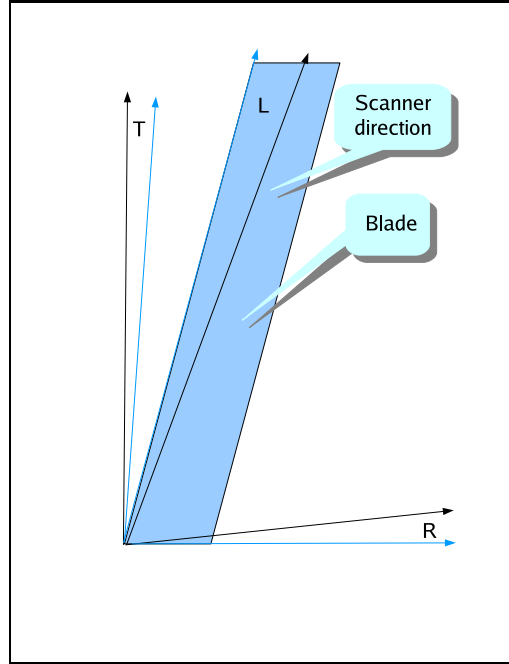


FIG. 1 – Positionnement de la lame dans le scanner, (L) angle de lacet, (R) angle de roulis et (T) angle de tangage.

Pour conclure sur le modèle à employer, il faut remarquer que les déformations 3D sont forcément faibles. En effet, les angles de roulis ou de tangage de la lame sont forcément petits de part le design du scanner. Dans le cas de dérèglement du plan du bras, l'erreur induite sur la géométrie sera faible, mais l'erreur de luminosité induite par la prise de vue confocale sera telle que les images seront très dégradées [MCGO-02] [PET-02]. Il convient dans ce cas de faire régler le scanner et de ne pas utiliser les images issues de telles mesures. Dans le cas où l'on soupçonnerait ce problème, il conviendrait de scanner une lame deux fois en la faisant pivoter horizontalement de 180 degrés entre deux prises de vue et de mesurer un éventuel gain de luminosité moyenne sur une même zone des deux images.

Notre but ici n'est pas de retrouver les paramètres de prises de vue mais de trouver une association entre les coordonnées dans le repère de l'image et les coordonnées dans le repère de la grille. Il nous suffit donc d'utiliser le fait que

toute combinaison d'applications affines en deux dimensions peut s'écrire sous la forme d'un système d'équations linéaires :

$$\begin{cases} X' &= a_{11}X + a_{21}Y + a_{31} \\ Y' &= a_{12}X + a_{22}Y + a_{32} \end{cases}$$

où X, Y sont les coordonnées d'un spot dans le repère de la grille, X', Y' dans le repère de l'image et les a_{ij} les coefficients du modèle. Celui-ci correspond le plus souvent à une combinaison rotation translation dans le plan de la lame, mais peut aussi intégrer toute déformation du positionnement de la lame du fait que l'on travaille en projection entre deux plans.

Théoriquement, ce système peut être entièrement déterminé par un triplet de points mais cela nous a semblé conduire à trop d'imprécisions. Les images de biopuces sont rectangulaires avec de fortes disparités entre les deux dimensions. Il est donc impossible de trouver une répartition suivant un triangle quasi isocèle des trois amers et donc les erreurs induites par un mauvais choix de pixels auront un effet non uniforme sur le positionnement de la grille. Le choix des points est effectivement délicat, car certaines zones peuvent ne pas comporter de spots exprimés et si l'opérateur groupe trop les points les erreurs induites dans le reste de l'image peuvent être démultipliées.

Par conséquent, nous utilisons une méthode de détermination des paramètres par moindres carrés utilisant au minimum quatre points, ce qui permet à l'opérateur de les répartir correctement, mais également d'absorber des petites erreurs de positionnement et de distinguer des incohérences flagrantes. En effet, la cohérence du choix des points peut se mesurer par la distance résiduelle restante entre les spots cliqués par l'opérateur et les positions calculées par le modèle. Pour que la correction soit acceptable, nous considérons que le maximum de cette erreur, appelé résidus, doit être inférieur à $\frac{1}{5}$ du rayon d'un spot.

Pratiquement, nous affichons l'image et la grille placée sur l'image par un simple centrage puis l'opérateur indique au moins quatre amers en cliquant sur l'image et sur la grille. Ensuite, nous calculons le modèle d'interpolation, affichons le résidus, et déplaçons la grille. L'opérateur peut alors déplacer les points jusqu'à obtenir une correction parfaite. Un résultat de ce traitement est montré sur la figure 2.

Cette méthode ne corrige bien évidemment pas les déformations locales, nous avons donc développé une méthode permettant de les compenser dont l'exposé fait l'objet de la partie suivante.

3.2 Correction des déformations locales de positionnement

La connaissance fine du positionnement des spots est cruciale pour la qualité des mesures en particulier quand les spots sont très rapprochés. En effet, après

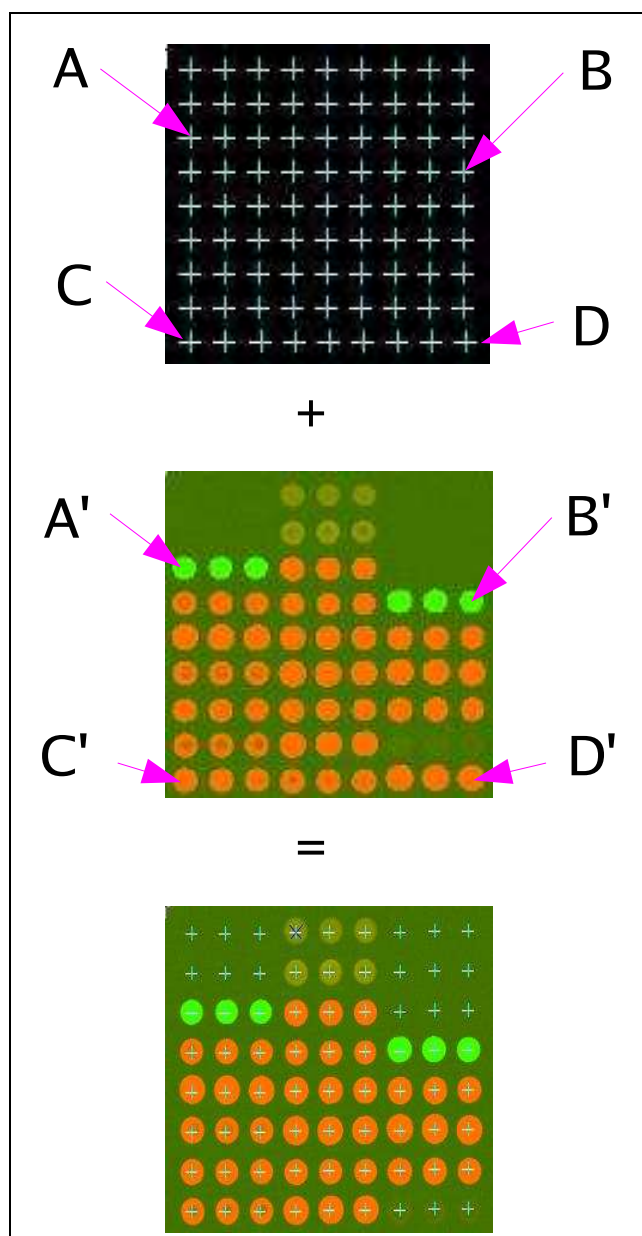


FIG. 2 – Positionnement global d'une grille à l'aide de quatre amers A,B,C,D.

le calage global de la grille, nous constatons souvent des erreurs résiduelles de valeurs suffisantes à provoquer la confusion entre une zone faiblement exprimée et le spot adjacent. Bien évidemment de telles erreurs sont ensuite éliminées, soit visuellement, soit par l'application de critères de qualité, mais l'effet de coupure des spots qu'elles provoquent sont gênantes sur aux moins deux aspects :

- elles empêchent d'utiliser le bord de la zone comme étant certainement du fond,
- elle risque de conduire à des spots coupés qui seront donc rejetés par les tests de qualité alors que corrects.

L'amélioration du positionnement est donc nécessaire pour obtenir une incertitude inférieure à la demi-distance interspot.

D'autre part, les spots étant très nombreux, il est impossible d'utiliser à grande échelle une méthode de calage assistée manuellement, il nous paraît donc souhaitable de développer une méthode de recherche de position totalement automatique. Ceci est possible à ce stade car la méthode de calage global est intervenue et nous permet donc de nous affranchir suffisamment des conditions expérimentales.

L'observation de ces déplacements [BALA-02] nous conduit à distinguer deux cas différents qu'une méthode de correction se doit donc de prendre en compte : le cas où le spot n'est pas à la place prévu mais sans être déformé, il est donc quasi circulaire, ou bien le cas où le spot a subi des déformations et éventuellement un déplacement. Sur une lame, ces deux types d'artefacts peuvent bien évidemment intervenir séparément ou ensemble. Le positionnement n'étant pas une phase de segmentation, il serait souhaitable d'utiliser des algorithmes indépendants de la forme et de l'expression du spot pour affiner la connaissance de leur position. Nous remarquons cependant que les spots bien exprimés et de forme quasi-circulaire fournissent une information fiable qu'il est peut être intéressant d'exploiter.

Nous avons donc développer ici une approche dual basée sur la corrélation avec un modèle circulaire associé à un calcul barycentrique basé sur les luminances. Le premier algorithme vise à se recalculer sur les spots qui n'ont que peu de déformations alors que le deuxième vise à prendre en charge les spots déformés et non uniformément exprimés.

Pratiquement, nous calculons la corrélation entre le premier canal image et un modèle circulaire binaire de même taille que les spots recherchés. Nous déplaçons ce modèle sur la fenêtre entourant le spot et calculons en chaque point la valeur de la corrélation. Nous sélectionnons alors sur ce corrélogramme la valeur maximum obtenue. Si celle-ci est supérieure à 0,75, nous considérons le point obtenu comme fiable. Nous réalisons ensuite le même calcul sur le deuxième canal ce qui nous donne un deuxième point sous les mêmes conditions. Cette technique permet de s'affranchir de la non expression éventuelle sur un des deux canaux. Nous calculons ensuite le barycentre de la zone entourant le spot. Pratiquement nous calculons le barycentre de la fenêtre entourant le centre du spot donné par la

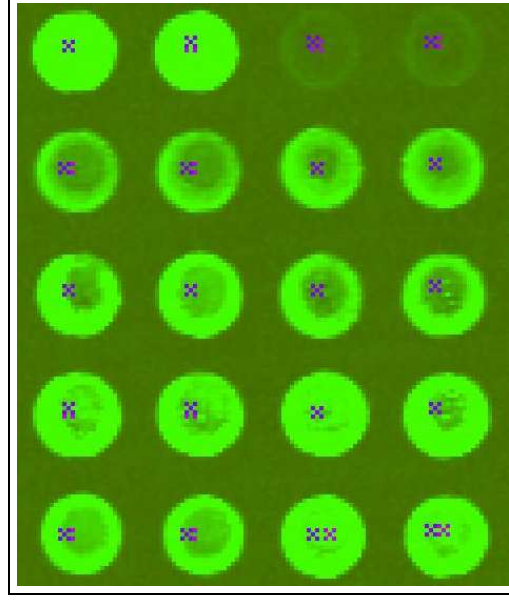


FIG. 3 – Position des barycentres d'intensité sur une image.

grille. À la fin de ce calcul, nous disposons de quatre pixels différents candidats à être le centre du spot. Ces informations sont fusionnées par la formule suivante :

$$\begin{cases} X &= \frac{\alpha_r X_{c_r} + \alpha_v X_{c_v} + \frac{1}{4}(X_{g_r} + X_{g_v})}{\alpha_r + \alpha_v + \frac{1}{4}} \\ Y &= \frac{\alpha_r Y_{c_r} + \alpha_v Y_{c_v} + \frac{1}{4}(Y_{g_r} + Y_{g_v})}{\alpha_r + \alpha_v + \frac{1}{4}} \end{cases}$$

où α_r et α_v sont les coefficients de corrélation linéaire obtenus, (X_{c_r}, Y_{c_r}) et (X_{c_v}, Y_{c_v}) les coordonnées des maxima de corrélation, (X_{g_r}, Y_{g_r}) et (X_{g_v}, Y_{g_v}) les coordonnées des centres de gravité respectifs au canal rouge et au canal vert. α_r et α_v sont bien évidemment seuillés à zéro s'ils sont inférieurs à 0.75.

Nous avons choisi une formule de fusion de ce type plutôt que de ne garder qu'un seul point, car sous réserve des seuils, ils peuvent tous être considérés comme fiables.

Ceci nous permet de lever les ambiguïtés telles que représentées sur la figure 4. Cette technique s'est avérée particulièrement efficace dans le cas de spots ayant la forme de donuts. À l'issue de ce traitement, nous disposons d'une grille affinée telle qu'elle est représentée sur la figure 4.

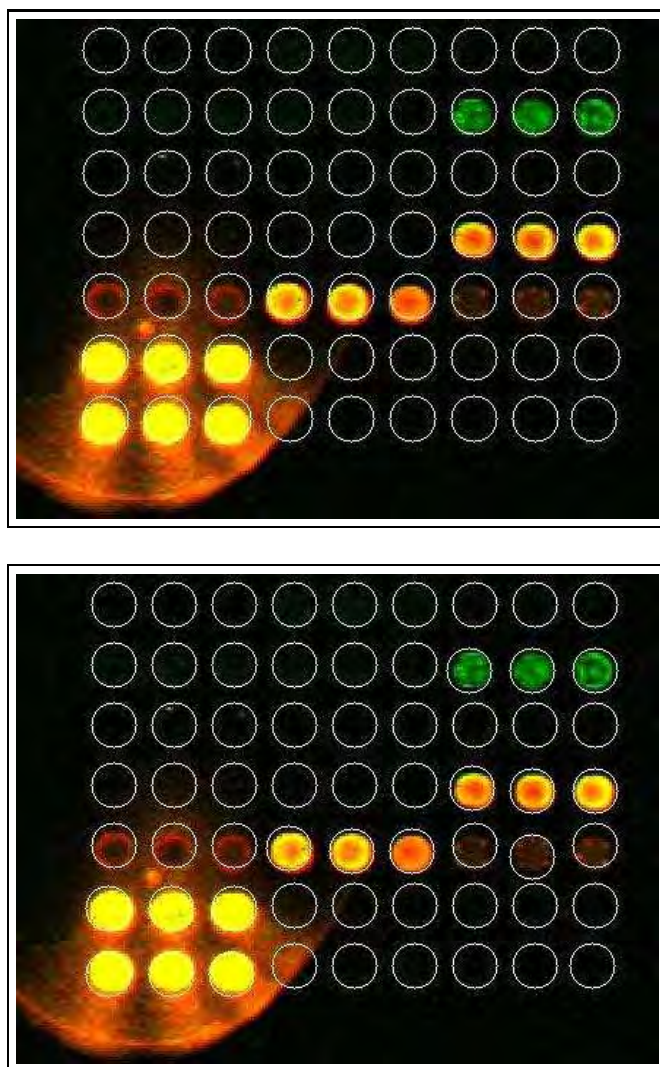


FIG. 4 – Affinage de la position des spots, en haut avant en bas après.

4 Segmentation

4.1 Estimation de la moyenne avant segmentation

Comme cela est dit dans la partie précédente, notre méthode de segmentation nécessite, pour son initialisation, une estimation de la moyenne du spot et du fond l'entourant. Cette estimation est problématique parce qu'on ne connaît pas les pixels constituant le spot et que cette connaissance est notre but. Une erreur importante sur ce paramètre remettrait évidemment en question la qualité de la segmentation.

Pour estimer cette moyenne, on utilise classiquement les pixels entourant le centre du spot donné par la grille ou une forme circulaire représentant la forme idéale du spot. Ces deux approches ont pour défauts d'être trop sensible au positionnement de la grille et d'augmenter les risques de confusion entre le fond et le spot. Dans le cas particulier de l'utilisation d'une forme, toute déformation des spots impliquera l'introduction de pixel de fond dans le calcul de la moyenne du spot et inversement. Ceci peut conduire dans des cas de faible contraste ou de grande erreur de positionnement, comme celle de la figure 5, à une segmentation totalement aberrante.

Dans le cas où nous chercherions à positionner une forme, l'erreur d'estimation du spot pourrait s'écrire sous la forme de la somme de l'erreur induite par le positionnement et de l'erreur induite par la déformation. Nous présentons donc ici une méthode qui permet de supprimer la nécessité de connaître finement la position du spot pour le calcul d'une valeur approchée des moyennes. La méthode des centiles décrite dans [DEMI-05] améliore un peu les estimations de moyenne mais reste trop conditionnée par la qualité géométrique des spots.

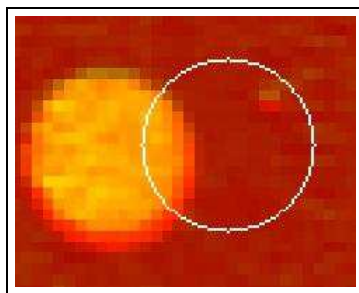


FIG. 5 – Erreur de positionnement d'une forme

En observant un grand lot d'image et en réalisant de nombreuses segmentations manuelles, nous nous sommes aperçus que si la forme des spots pouvait être très variable, leur surface varie peu. En fait, les déformations de spots sont la plu-

part du temps due à un glissement de produit sur la lame ce qui ne change pas sur l'image le nombre de pixels qui les constituent. Notre méthode d'estimation est donc basée sur ce fait.

Nous procédons à l'estimation de la moyenne du spot en observant son influence sur la moyenne du fond entourant le spot. Pratiquement, cela consiste à écrire que la moyenne d'une zone (Z) contenant un spot est la somme pondérée de la moyenne des valeurs du fond (F) et des pixels (S) comme cela apparaît sur la figure 6.

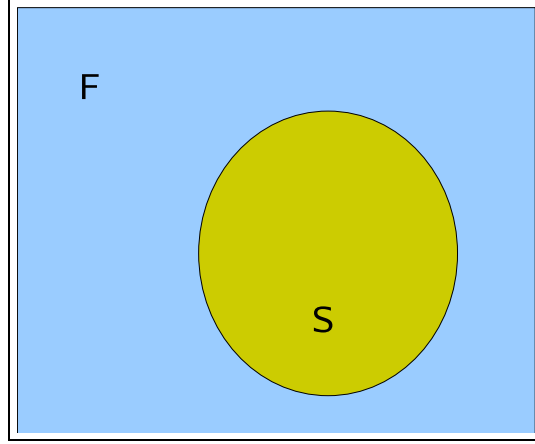


FIG. 6 – Boîte Z entourant un spot S avec un fond F.

La moyenne de la zone Z ($E(Z)$) s'écrit :

$$E(Z) = \frac{\sum_{x,y \in F} P(x,y) + \sum_{x,y \in S} P(x,y)}{\text{card}(F) + \text{card}(S)}$$

ce qui nous permet de déduire la moyenne $E(S)$ de la zone S en fonction de $E(Z)$ et la moyenne du fond ($E(F)$) :

$$E(S) = \frac{E(Z)\text{card}(Z) - E(F)\text{card}(F)}{\text{card}(S)}$$

La moyenne ($E(F)$) du fond est mesurée sur le bord de la boîte entourant le spot. La surface théorique du spot ($\text{card}(S)$) est donnée par les caractéristiques du spoteur et par la résolution choisie sur le scanner. La population du fond ($\text{card}(F)$) est donnée par $\text{card}(Z) = \text{card}(F) + \text{card}(S)$.

Afin de démontrer l'intérêt de cette méthode, prenons l'exemple d'une erreur de positionnement du centre du spot de 0 à $\frac{1}{2}$ rayon pour un spot de rayon 11 pixels. Le tableau de la figure 7 nous donne les erreurs de surface avec de tels

Décalage en pixel	erreur en pixel	erreur % de la surface
0	0.0	0.0
1	23.0	6.10
2	44.0	11.67
3	65.0	17.24
4	86.0	22.81
5	107.0	28.38

FIG. 7 – Erreur d’initialisation sur un cercle de 11 pixels.

parametres et les courbes de la figure 8 les erreurs de moyenne induites pour différents niveaux de contraste.

Nous remarquons sur ces figures que les erreurs de surface sont très importantes et induisent des erreurs de moyenne qui dépendent fortement du contraste local de la zone étudiée. Il est notable que l’erreur induite sur la moyenne est de l’ordre des contrastes que nous cherchons à mesurer et justifie donc totalement l’emploi de notre méthode d’estimation.

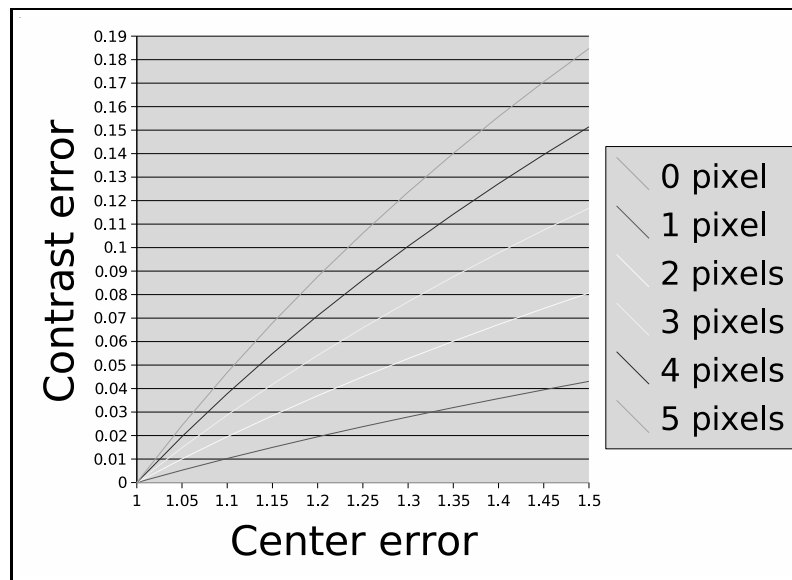


FIG. 8 – Erreur d’estimation de la moyenne d’un spot pour des contrastes allant de 5 % à 50 % et pour des décalage de 1 à 5 pixels.

Par contre, l’erreur due aux variations de surfaces induites par l’arrachement de matière du spot perdue dans les deux approches d’où la nécessité de recourir à

une étape de segmentation dont le but est justement de calculer la meilleure forme possible pour le spot.

4.2 Que faire avec deux résultats de segmentation

Avant même de segmenter les images, il est impératif de poser les conditions d'emploi des deux canaux à notre disposition. En effet, en présence de deux images nous pouvons segmenter en nous appuyant directement sur les deux, ou choisir une stratégie de segmentation séparée puis de fusion des résultats.

La première stratégie ne doit être employée que quand les deux informations tirées des deux images participent à la définition d'une même forme, ce qui n'est pas le cas dans le problème qui nous préoccupe. En effet, les biopuces sont conçues pour que l'information donnée par les deux images soit différente dans leur signification biologique et les sondes déposées sont hybridées avec deux ARN différents dans le but de donner des fluorescences différentes. En théorie, on doit obtenir une même forme mais avec des intensités très différentes. Ces intensités peuvent être très proches ou très corrélées mais ne le sont pas dans le cas où le niveau d'expression d'un seul canal est très faible. En pratique, il nous faut distinguer trois cas :

1. un canal bien exprimé et l'autre pas du tout
2. deux canaux bien exprimés de même forme
3. un canal bien exprimé de forme très proche du dépôt et un autre également bien exprimé mais ayant une forme très éloignée du dépôt.

Des exemples de ces trois situations sont repris sur la figure 9.

Pour statuer sur le premier cas et sur le dernier, il faut en revenir à la physique de la formation des images qui nous intéressent ici. Quand un spot s'exprime, cela signifie que la lumière perçue par les capteurs provient de la fluorescence et que la réponse du fond n'est plus visible. A l'inverse, dans une zone non hybridée, la lumière mesurée par le scanner provient uniquement du fond. Dans ce dernier cas, la sonde est transparente et ne fournit alors aucune information de forme exploitable et cela même si une certaine déformation du signal de fond peut être mesurée. Il convient donc d'adopter le schéma de fusion donnée suivant :

- Dans le cas numéro un, la zone segmentée comme appartenant à la sonde doit provenir uniquement de l'image bien exprimée,
- Dans le cas numéro deux, la zone d'expression est celle où les deux segmentations ont conclu à la classe spot,
- Dans le cas numéro trois, l'information du canal qui a une forme très différente du dépôt doit être rejetée ou sursegmentée vers celle de l'image bien exprimée. Il faut noter que ce dernier cas est très rare.

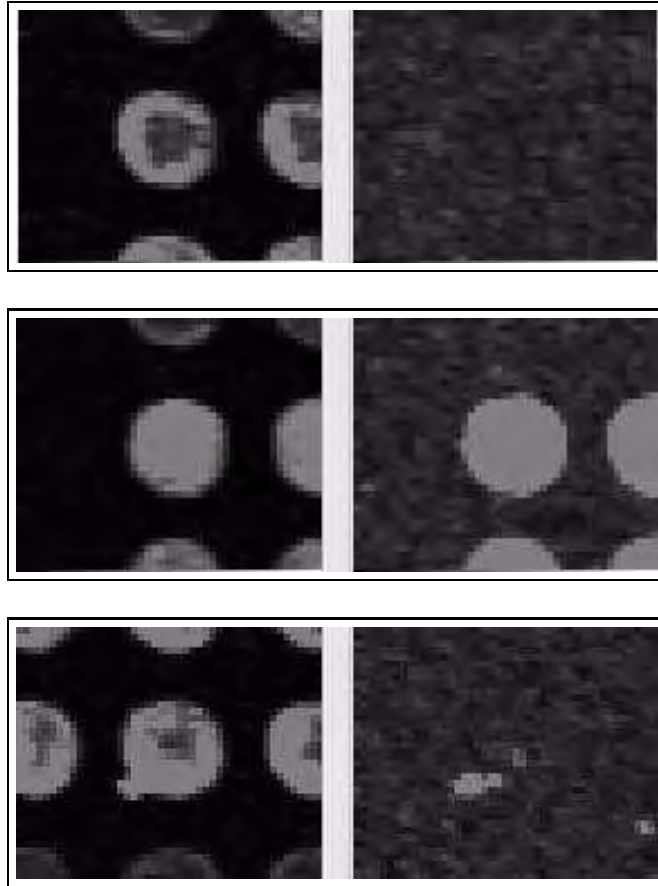


FIG. 9 – Trois situations d’expresion d’images de biopuces, à gauche le canal rouge (cy5) à droite le canal vert (cy3).

Ceci étant posé, nous avons choisi de segmenter séparément les images des deux canaux CY5 et CY3 puis d'appliquer le schéma de fusion défini ci-dessus en distinguant les cas par une post-estimation des moyennes. Concrètement, nous calculons les moyennes du fond et du spot, après segmentation sur chaque canal, dans l'image d'un canal en utilisant la forme segmentée sur l'autre canal et comparons les variations de statistiques sur l'un et l'autre. Ceci nous permet de distinguer dans quel cas nous nous trouvons et d'orienter notre fusion. Les parties suivantes sont consacrées à la définition de notre algorithme de segmentation.

4.3 Pourquoi utiliser un champ de Markov

Notre segmentation est basée sur le fait que les statistiques du premier ordre des spots sont différentes de celles du fond. Les biologistes sont intéressés par la médiane (Ralph) des pixels constituant les spots car ils considèrent celle-ci comme plus représentative de l'expression des gènes que la moyenne. Notre segmentation est cependant basée sur la moyenne pour au moins deux raisons :

- si nous segmentons grâce à la médiane, nous risquons de biaiser la mesure finale
- la moyenne est beaucoup plus rapide à calculer que la médiane qui demande de plus une réévaluation complète quand on ajoute ou retire un individu de l'espace d'étude. Comme les images comportent de nombreux spots, cette rapidité est essentielle pour pouvoir exploiter notre travail dans un logiciel interactif tournant sur des ordinateurs personnels.

Nous utilisons donc une classification itérative basée sur la distance d'un pixel à la moyenne de la classe à laquelle il est censé appartenir. Dans ce cadre, une première approche très classique [PAL-93] qui s'appelle une classification par nuée dynamique peut être tentée. Concrètement, nous calculons dans ce cas pour chaque pixel la distance à la moyenne de chaque classe et en déduisons son appartenance. La distance utilisée est alors de la forme :

$$D(i, c) = |X - E(X_c)|$$

où $E(X_c)$ est la moyenne de la classe des pixels de la classe C. Nous appelons D attache aux données dans la suite.

Une fois tous les pixels classés, nous réévaluons la moyenne des classes et reclassons chaque pixel. Ce procédé est réalisé localement pour chaque sonde sur une petite fenêtre entourant le spot. En effet, les valeurs d'expression étant très disparates d'un spot à l'autre, leur moyenne est très différente et donc une approche globale à l'image serait aberrante. Bien évidemment, l'utilisation de cette méthode requiert une très bonne estimation de la moyenne des spots et du fond. Le résultat que nous obtenons est une image de masques, superposable à l'image

d'origine, dont les pixels ont une valeur zéro pour le fond et une valeur correspondant au numéro du spot. Une fois toutes les zones de l'image traitées et agrégées, nous obtenons une localisation de chaque spot.

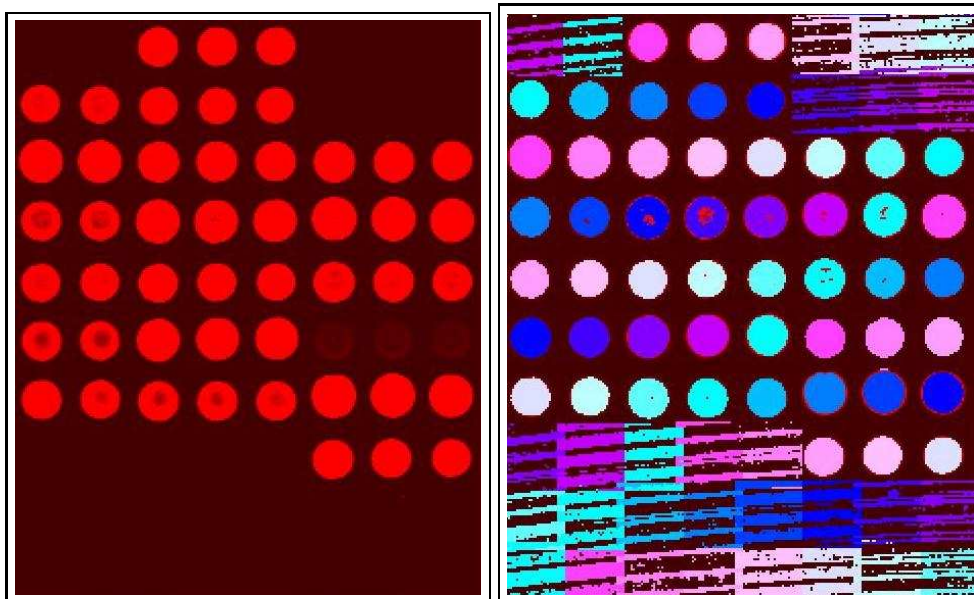


FIG. 10 – Classification par nuées dynamiques : à gauche, le canal rouge d'une image de biopuce, à droite, le résultat.

Sur la figure 10 nous montrons le résultat de ce traitement et nous pouvons constater qu'il est loin d'être parfait. Nous remarquons en particulier que certains spots, au centre de l'image, comportent des trous et que le traitement des zones de faible intensité aboutit à un résultat aberrant.

Pour améliorer cela, il convient d'introduire des corrections basées sur l'analyse des défauts de ce type de classification. L'objet de la partie suivante est l'exposé de notre schéma de classification markovien répondant à cette problématique.

4.4 Notre schéma de classification Markovien

Les classifications telles que décrites ci-dessus n'utilisent qu'un critère basé sur la valeur des pixels. Ceci explique pourquoi beaucoup de pixels isolés sont mal classés comme on peut le voir sur la figure 10. Le problème vient du fait qu'il existe dans l'histogramme de l'image une zone confondue où se mélange des pixels appartenant au fond et aux spots. Cette confusion, telle qu'elle est représentée sur la figure 11 est une limitation connue des classifications par nuée dynamique et elle ne peut être compensée que par l'introduction d'informations supplémentaires.

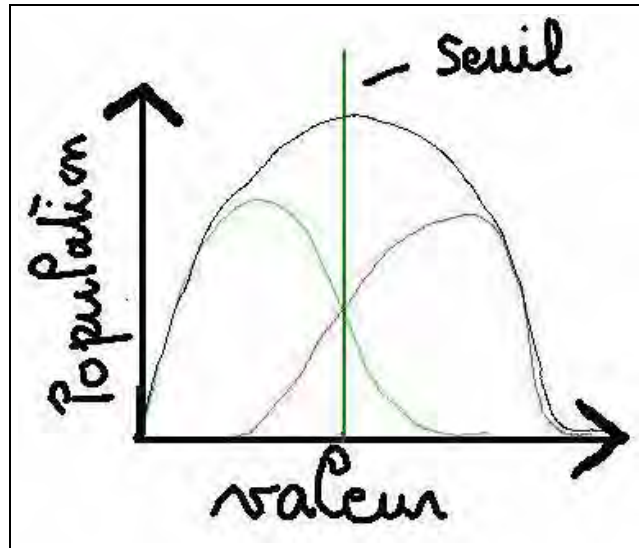


FIG. 11 — Sur ce schéma, nous voyons l'histogramme cumulé de deux populations avec une partie indécidable au milieu.

De ce fait, nous allons introduire dans notre classification des critères spatiaux qui nous permettront de mieux classer les pixels isolés ayant une valeur aberrante. Les spots étant le plus souvent connexe par arc, il apparaît qu'un pixel a forcément de nombreux voisins de la même classe que lui. Il est donc possible d'enrichir l'information de distance à l'aide du comptage des voisins de même classe. Nous avons donc modifié le calcul de la fonction d'appartenance à une classe de façon à relâcher l'attache aux données au profit des critères spatiaux et cela pour les pixels dont l'intensité ne permet pas de les affecter de façon sûre. Ceci revient à modifier le calcul de la distance des classes pour mieux gérer la partie incertaine, c'est-à-dire les pixels dont les valeurs sont comprises entre la moyenne de la classe "fond" et celles de la classe "spot".

D'autre part, la distance à une des deux classes peut se voir comme l'inverse de la possibilité d'appartenir à l'autre classe. Cette notion est difficile à manipuler du fait des non linéarités qu'elle engendre et de ces passages par la valeur zéro. Ceci explique pourquoi dans la suite nous allons exprimer le facteur d'appartenance à une classe comme un potentiel. Celui-ci comporte deux termes, un exprimant l'appartenance à une classe et l'autre l'appartenance à l'autre classe. La soustraction de ces deux éléments nous donnera une valeur positive ou négative exprimant la possibilité d'appartenir à l'une ou l'autre classe.

Soit $I(x, y)$ l'intensité du pixel de position x, y , nous écrivons le potentiel d'appartenir à la classe c relativement à la classe c' :

$$\begin{aligned}
P(x, y, c, c') &= \alpha(I(x, y), c) \\
&- \alpha(I(x, y), c') \\
&+ e(x, y, c) \\
&- e(x, y, c')
\end{aligned}$$

où $\alpha(I, c)$ représente le potentiel d'appartenance à la classe c pour l'intensité I et $e(x, y, c)$ l'énergie potentielle fournie par le voisinage. La difficulté de mise au point d'une telle classification réside dans le choix de ces termes et dans leur équilibrage qui est toujours délicat.

Pour la classe "fond" les contraintes sont :

- la nature du bruit fond, composé de chatoiement et de fragments de sonde hybridés et la connexité des spots, qui induit qu'ils sont sous-représentés en surface, nous a conduit à choisir des potentiels symétriques. En effet, nos objectifs pour la classe fond sont d'éliminer des pixels isolés de fortes valeurs classés à tort comme appartenant à un spot ce qui suppose d'utiliser un critère spatial fort. Ceci aura pour défaut de renforcer la zone la plus représentée et il nous faudra compenser cette effet par ailleurs.
- nous souhaitons que, pour des zones de moyenne proche de celle du fond mais incluse dans le spot, la classification les considère comme appartenant aux spots. Ceci implique une attache aux données forte autour de la moyenne du fond mais qui décroît rapidement.

Pour la classe "spot", les contraintes sont :

- les valeurs proches de la moyenne doivent être affectées à la classe "spot" sous réserve de voisinage.
- les pixels proches de la moyenne du fond doivent être rejetés sauf en frontière de façon à corriger les effets de sur-représentation surfacique.
- l'attache aux données doit croître au fur et à mesure qu'on s'éloigne de la moyenne du fond et que l'on s'approche la moyenne du spot.

À partir de ces deux jeux de contraintes, nous avons fabriqué le potentiel linéaire par morceaux $\alpha(I, c)$ défini par les formules suivantes :

si $c = F$

$$\begin{cases} I < \overline{F} & \alpha(I, F) = 1 \\ \overline{F} < I < \frac{\overline{F} + \overline{S}}{2} & \alpha(I, F) = \frac{3}{8} \left(\frac{\frac{\overline{F} + \overline{S}}{2} - I}{\overline{S} - \overline{F}} \right) + \frac{1}{4} \\ \frac{\overline{F} + \overline{S}}{2} < I & \alpha(I, F) = 0 \end{cases}$$

si $c = S$

$$\begin{cases} I < \overline{F} & \alpha(I, S) = 0 \\ \overline{F} < I < \frac{\overline{F} + \overline{S}}{2} & \alpha(I, S) = \frac{1}{4} \\ \frac{\overline{F} + \overline{S}}{2} < I < \overline{S} & \alpha(I, F) = \frac{3}{8} \left(\frac{I - \overline{F} + \overline{S}}{\overline{S} - \overline{F}} \right) + \frac{1}{4} \\ \overline{S} < I & \alpha(I, F) = 1 \end{cases}$$

où \overline{S} et \overline{F} sont les moyennes estimées du spot et du fond. Ce potentiel α est représenté graphiquement en fonction de l'intensité I sur la figure 12.

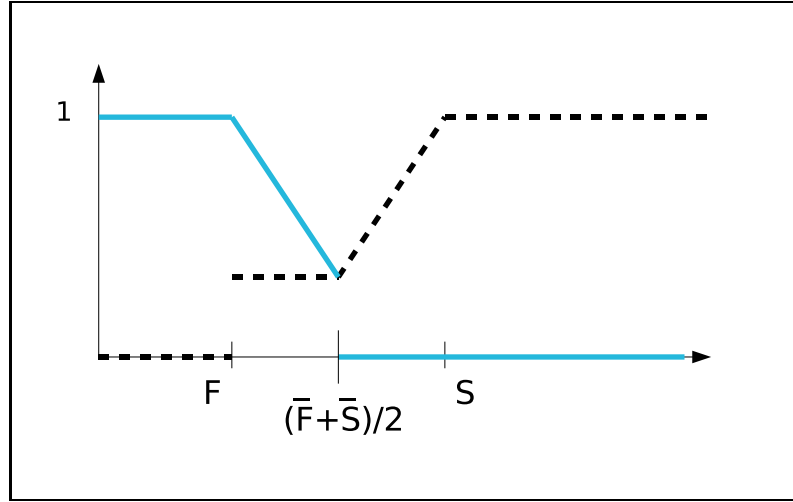


FIG. 12 – Sur cette figure nous voyons le potentiel α pour la classe "fond" en clair et pour la classe "spot" en tireté.

Le terme $e(x, y, c)$ est obtenu en comptant le nombre de voisins de la classe c se trouvant dans un voisinage en huit connexités, c'est-à-dire limité aux huit pixels entourant le spot. Ce choix est le résultat de divers tests et permet de très bons résultats en limitant au maximum les calculs. Avec cet ensemble de contraintes, il faut maintenant fabriquer l'image de classe c minimisant la fonctionnelle P en tout point de l'image original. Pour cela certaines propriétés de notre fonctionnelle sont précieuses et notamment le fait que le processus de régularisation spatiale en huit connexités nous permet d'affirmer que nous sommes dans un cadre markovien [DESC-93].

Cela nous indique que le problème d'optimisation a une solution et que par conséquent, un schéma itératif de recherche par minimisations successives convergera vers cette solution.

Cela implique également qu'il est possible de segmenter l'image spot par spot sans changer le résultat. Nous pouvons donc découper arbitrairement en morceaux

l'image sans changer l'aspect de la solution. Notons également que ce découpage est conforme à l'esprit des biopuces où chaque sonde est en réalité une expérience indépendante.

Le problème principal que pose, par contre, un tel schéma est sa complexité. Une recherche directe du minimum absolu serait beaucoup trop longue en particulier dans le cadre d'un logiciel interactif. Nous procèderons donc ici par recuit simulé. Cette technique, sous hypothèse markovienne en particulier, permet de garantir la convergence vers une solution proche de la solution réelle en ayant le choix d'arrêter le procédé suivant certains critères simples.

Notre algorithme se déroule donc en trois phases : une phase d'initialisation, une phase de classification itérée et une phase de nettoyage éventuel du résultat.

La phase d'initialisation consiste à :

- découper l'image en sous-images autour des spots
- estimer par la méthode décrite dans la partie 4.1 la moyenne du fond et du spot pour chaque image
- pour chaque site, on tire un masque de classe suivant un processus poissonnien ⁶
- pour chaque image, nous constituons une liste aléatoire d'ordres de traitement des pixels.

La phase de classification consiste pour chaque image à :

- prendre un pixel,
- calculer le potentiel pour la classe courante donnée par l'image de classe.
- Si ce potentiel est négatif, on change la classe, sinon on calcule l'expression $x = \log(\frac{P}{t})$ et on tire aléatoirement un nombre x' suivant une loi uniforme entre 0 et 1. Si $x' > x$, nous changeons la classe du pixel bien que le potentiel indique que le pixel est plutôt bien classé. Ce choix d'opposition constitue le coeur de l'algorithme de recuit simulé et permet de ressortir des minima locaux de potentiel sans devoir essayer toutes les combinaisons possibles.
- Nous passons ensuite aux pixels suivants jusqu'à épuiser la liste de traitement.
- A ce moment, nous évaluons le nombre de pixels qui ont été modifiés et si celui-ci est supérieur à 1 pour 1000, en pratique supérieur à zéro compte-tenu de la taille des spots couramment employés, nous réitérons le procédé avec $t_n = 0.95 * t_{n+1}$.
- une fois la convergence atteinte, nous passons à l'image suivante.

La phase de nettoyage, parfois appelé trempe, consiste à attribuer les éventuels pixels isolés à la classe qui les entoure par simple comptage.

⁶c'est-à-dire une image comprenant des pixels de valeur correspondant à une classe recouvrant l'image à traiter.

Ces algorithmes demandent beaucoup de calculs mais les propriétés des champs de Markov nous permettent de parallélisés sans risque le calcul des différentes imagerie. Ceci permet de raccourcir significativement la durée du traitement sur les machines dualcore ou multiprocesseurs. Le tableau de la figure 13 donne quelques temps de calcul indicatifs réalisés sur une machine de type PC ayant un processeur à 3 GHz.

Image size	Spot size	nb spot	time image
980x761	11 p	324	129 s
2200x7494	11 p	2267	283 s

FIG. 13 – Tableau des temps CPU nécessaires au calcul de la segmentation des biopuces par recuit simulé.

La partie suivante est consacrée à l’analyse des résultats de ce traitement sur une série d’images de biopuces.

4.5 Résultat de la segmentation Markovienne

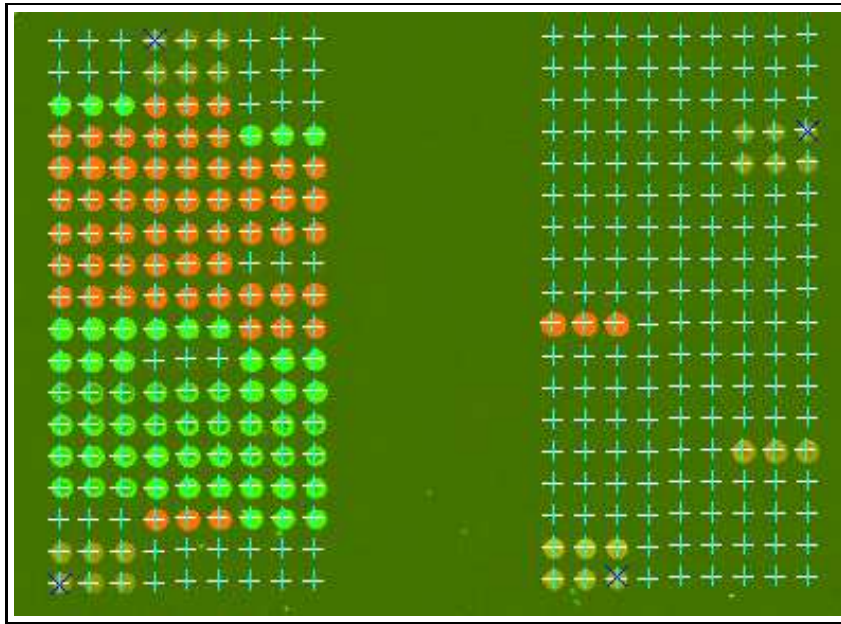


FIG. 14 – Image simple de biopuces.

Nous ferons la présentation de nos résultats sur deux séries d’images, une série facile, telle celle de la figure 14, et une série plus difficile, telle celle de la figure

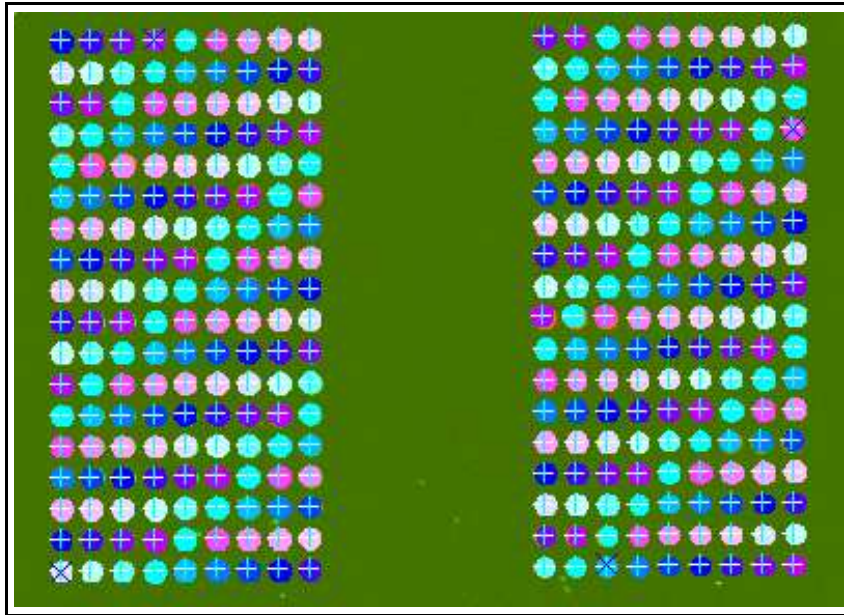


FIG. 15 – Image de la figure 14 segmentée à l'aide la technique du cercle fixe.

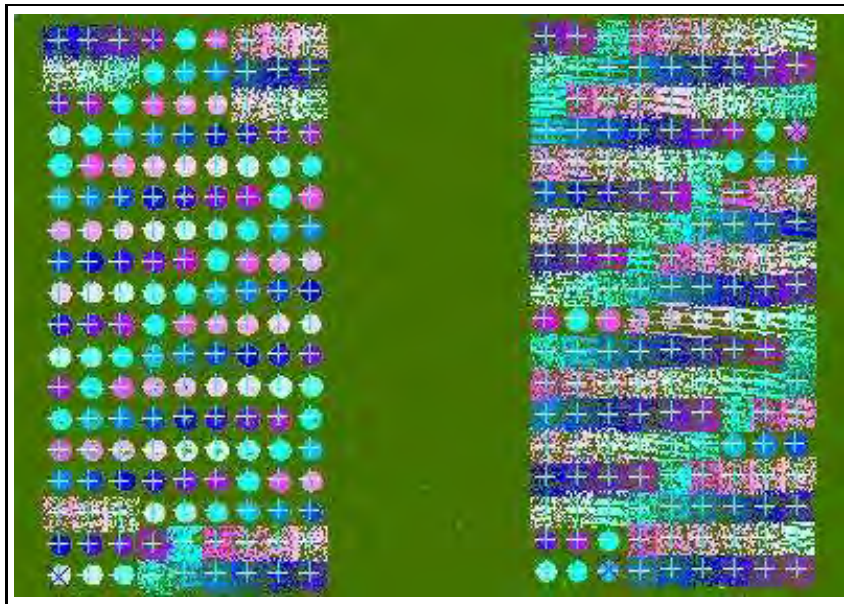


FIG. 16 – Image de la figure 14 segmentée à l'aide d'une classification par nuées dynamiques.

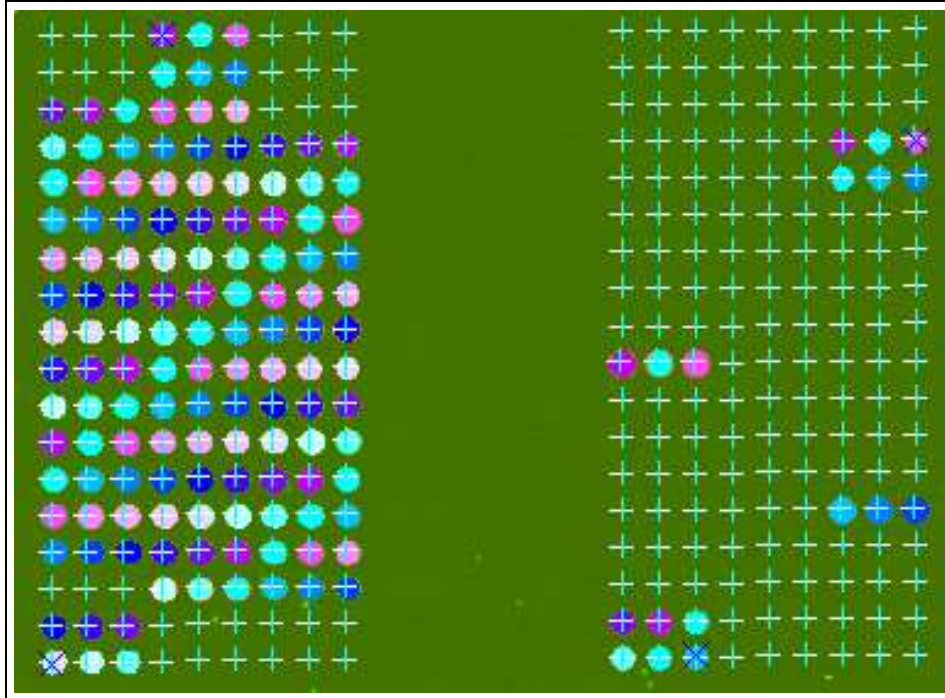


FIG. 17 – Image de la figure 14 segmentée à l’aide d’une classification par notre technique.

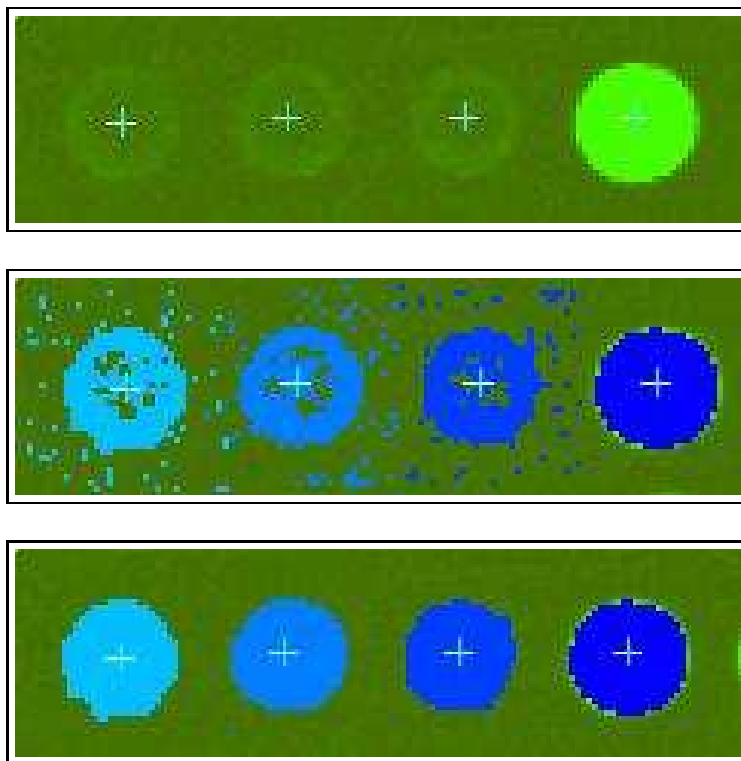


FIG. 18 – En haut, un extrait de l’image de la figure 14, au milieu, la segmentation par nuées dynamiques et en bas grâce à notre technique.

19 pour montrer les avantages et les limites de notre approche. Comme cela est dit précédemment, une bonne segmentation doit comporter au moins trois qualités :

- ne pas segmenter les spots non hybridés,
- se comporter correctement en présence d’artefacts,
- segmenter au plus juste les spots hybridés, c’est-à-dire donner le taux de pixels de fonds détectés comme appartenant à une sonde le plus faible possible.

Pour juger de la première qualité, nous ne pouvons pas considérer directement le nombre de spots hybridés détectés. En effet celui-ci dépend du nombre de spots non hybridés qui dépend de la biologie d’expérience et non de l’acquisition d’images. Il convient donc de calculer la probabilité de se tromper qu’a un algorithme quand il détecte un spot. Dans l’exemple de la figure 14, 100 % des spots détectés le sont correctement, c’est-à-dire qu’aucun spot non hybridé n’est détecté. Sur une base d’images plus vaste, le taux de détection que nous obtenons est alors de 99 % des bons spots et 99,9 % des spots non hybridés ne sont pas segmentés. Ces résultats sont bien évidemment meilleurs que ceux des cercles fixes qui ne sont pas prévus pour gérer ce genre de cas. Les nuées dynamiques atteignent un taux d’erreur voisin de celui des cercles fixes. Ce résultat peut paraître excellent, mais il ne faut pas oublier que cela implique plusieurs erreurs sur une lame à 10 000 sondes.

Le comportement en présence artefact ne peut se juger que visuellement car il n’existe pas de recensement exhaustif des défauts possibles dans ce type d’images. Les deux types les plus gênants sont liés au bruit de fond et aux arrachements de matière sur les sondes qui produisent des zones très brillantes après hybridation. La figure 18 présente un résultat de comportements face au bruit de fond où nous constatons que notre technique est beaucoup plus fiable que les nuées dynamiques. Sur la figure 21 nous présentons le résultat de notre traitement dans le cas d’une image plus difficile (figure 19). Sur ces données, nous voyons plusieurs cas de pollution de sonde provenant d’arrachements d’autres spots avec différentes tailles et proximités de la sonde polluée. Nous constatons alors que le comportement de notre méthode est globalement satisfaisant, car les zones de pollution ne sont pas segmentées dès qu’elles atteignent une taille significative. Une sonde, située sur la deuxième ligne en partant du bas en quatrième colonne, est rejetée car trop polluée, alors que sa voisine qui l’est significativement moins est simplement sous-segmentée. Là encore, la comparaison avec les nuées dynamiques est très en notre faveur notamment en ce qui concerne l’homogénéité des masques produits.

Pour ce qui est de juger de l’aptitude d’une segmentation à intégrer un minimum de pixels de fond dans les spots produits, la seule méthodologie valable consiste à tracer un grand nombre de spots à la main puis à calculer les matrices de confusion entre les segmentations et le masque manuel [KEND-52]. Cette opération est longue et fastidieuse si nous voulons traiter un grand nombre de

cas. Nous avons donc procédé en tirant au sort 20 spots dans 32 lames différentes ce qui permet de balayer un grand nombre de cas sans dépenser trop de temps. Parmi ces spots, 15 par puce ont été utilisés pour calculer la matrice de confusion et cinq pour vérifier sa robustesse. Nous avons écarté trois lames du fait de l'instabilité des résultats et nous obtenons donc la matrice suivante :

	Cercle fixe	Nuées dyna- miques	Markov
Manuel	81 %	79 %	97 %

Sur ses résultats, nous remarquons que notre méthode produit des masques tout à fait convenables, bien meilleurs que les deux autres méthodes mêmes si les résultats des cercles fixes peuvent être très bons dans le cas de puces d'excellente qualité. Notre méthode a donc les trois qualités qui correspondent à notre cahier des charges. Nous remarquons de plus que l'adaptation de notre segmentation aux petites déformations de grille, non compensées parce les deux techniques de calages présentées précédemment, est excellente. Ceci montre bien la supériorité des méthodes à base de classification pour la détection de la forme des spots sur les images de biopuces.

5 Conclusion et perspectives

Le travail présenté ici atteint les buts que nous nous sommes fixés, c'est-à-dire :

- traiter des biopuces provenant de constructeurs quelconques
- segmenter de façon robuste les forts et faibles contrastes
- tolérer l'utilisation de lames de qualité imparfaite

Notre méthode exploite au mieux les propriétés des différentes données disponibles (grille, images,...) et elle utilise au minimum les interventions d'opérateurs. Comme annoncé, nous avons mis au point une méthode rapide de calage de grille et une méthode de segmentation originale donnant des résultats satisfaisants. Ce travail est implémenté dans un logiciel interactif disponible à partir de la page [http ://gargamel.isima.fr/~gouinaud/](http://gargamel.isima.fr/~gouinaud/) dont un manuel est en cours de rédaction.

Notre approche va être poursuivie suivant deux axes visant toujours l'amélioration de la qualité des mesures d'expression. Nous souhaitons mettre au point un critère de qualité des formes obtenues par une caractérisation géométrique des masques et mettre au point des critères de qualité des statistiques obtenues à partir des valeurs segmentées. Un des buts est de régler les problèmes tels que des effets *donuts* (spot à trous) sans devoir recourir à une nouvelle phase de segmentation.

Dans un autre aspect, nous travaillons sur les techniques de calibrage rouge vert afin d'améliorer cette fois la qualité de comparaison entre deux populations

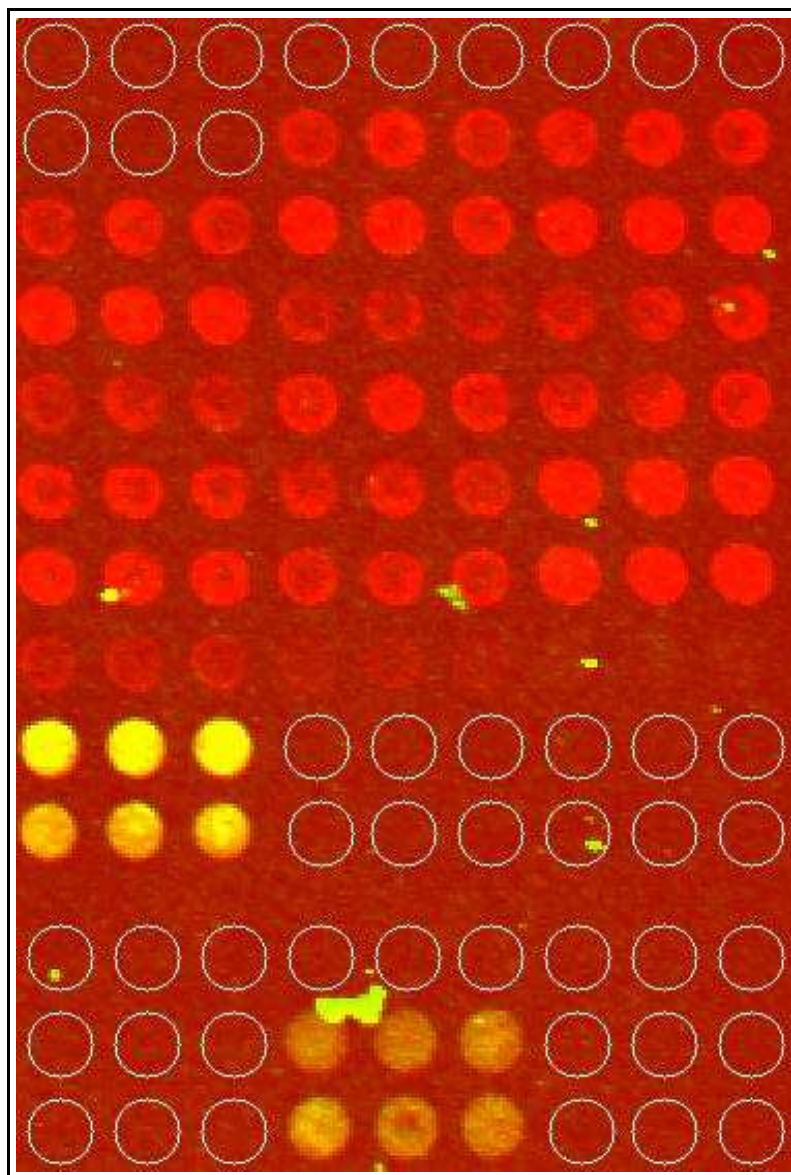


FIG. 19 – Résultat brut de notre traitement sur une zone d'image contenant de nombreux artefacts, les cercles sont sur les zones vides par production de la puce.

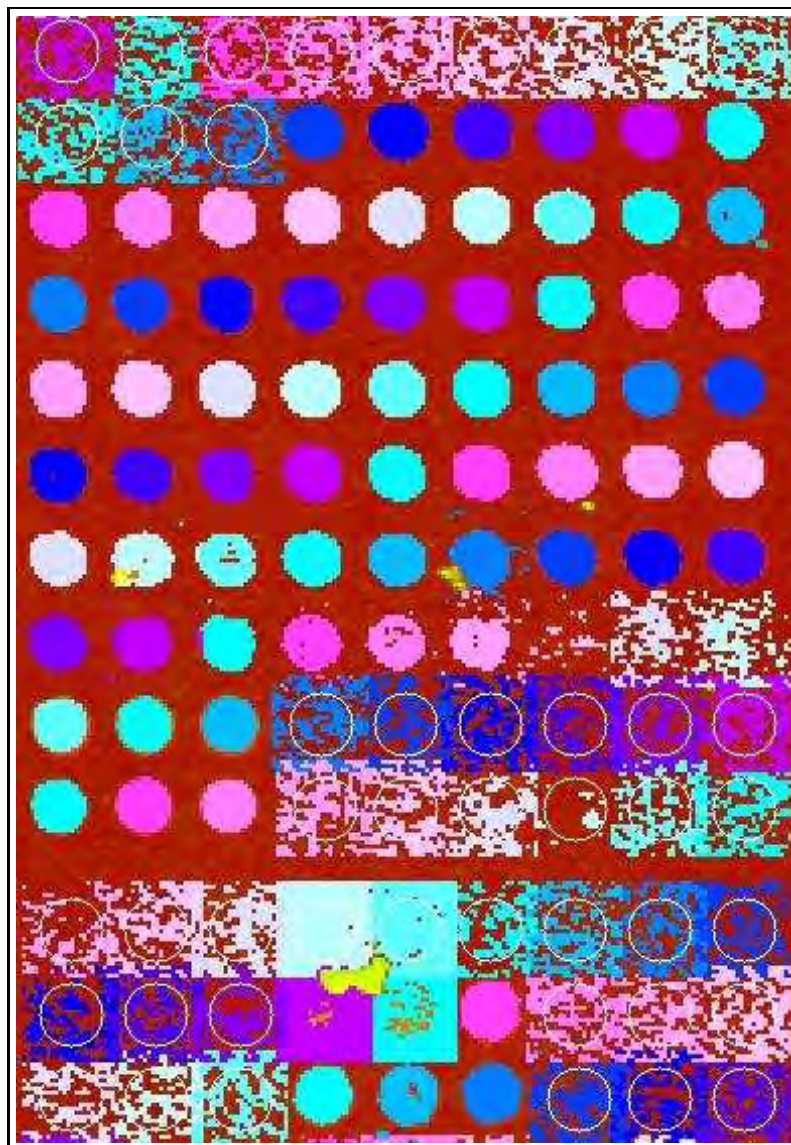


FIG. 20 – Résultat d'une classification par nuées dynamiques.

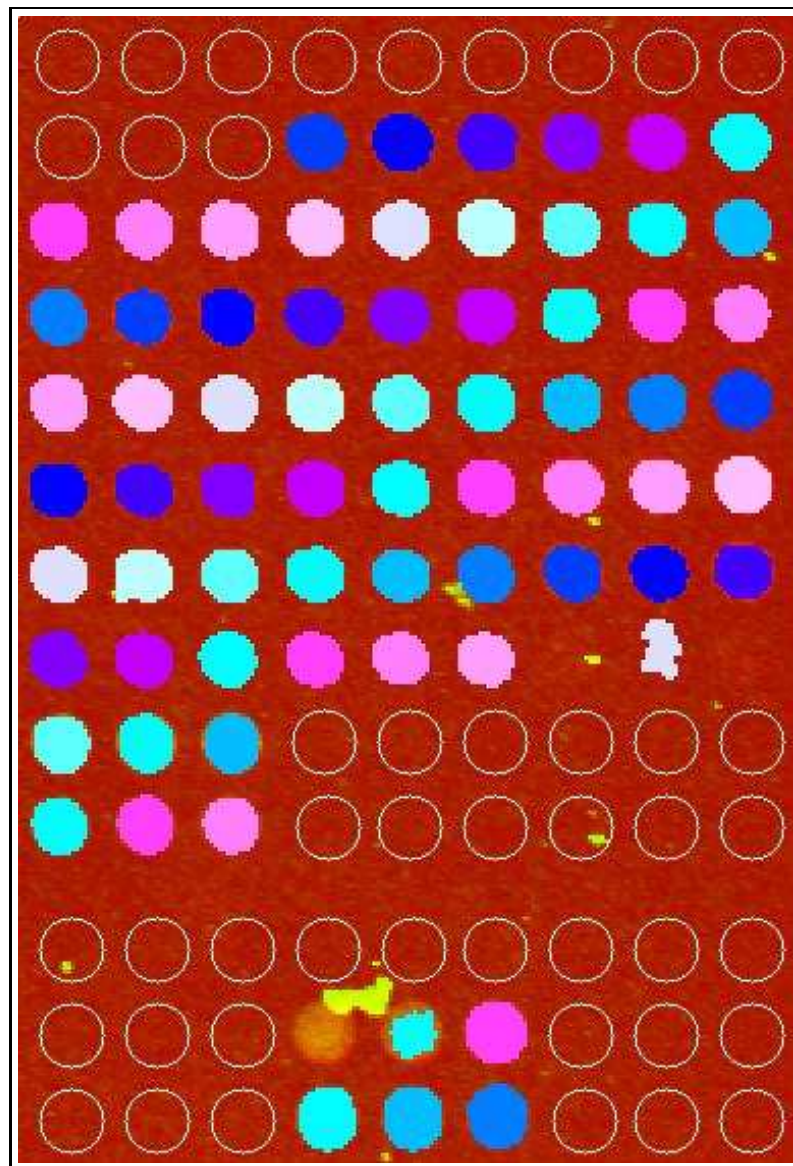


FIG. 21 – Résultat de notre traitement sur l'image de la figure 19.

de gène qui est souvent le but ultime de l'utilisation de biopuces.

Références

- [BALA-02] Y. Balagurunathan, E.R. Dougherty, and Yidong Chen. Simulation of cDNA microarrays via a parameterized random signal model. *Journal of Biomedical Optics*, 7(3) :507–523, July 2002.
- [BAR-06] Vincent Barra. Robust segmentation and analysis of DNA microarray spots using an adaptive split and merge algorithm. *Computer Methods and Programs in Biomedicine*, 81(2) :174–180, February 2006.
- [BROW-01] C. S. Brown, P.C. Goodwin, and P. K. Sorger. Image metrics in the statistical analysis of DNA microarray data. *PNAS*, 96(16) :8944–8949, July 2001.
- [DEMI-05] Omer Demirkaya, Musa H. Asyali, and Mohamed M. Shoukri. Segmentation of cDNA microarray spots using Markov random field modeling. *Bioinformatics*, 21(13) :2994–3000, 2005.
- [DESC-93] X. Descombes. *Champs markoviens en analyse d'images*. 93 E 026, ENST, Paris-France, 1993.
- [EISE-99] Michael B. Eisen and Patrick O. Brown. DNA arrays for analysis of gene expression. *Methods in Enzymology*, 303 :179–205, 1999.
- [GIUL-07] Giuliano Antoniol and Michele Ceccarelli. Microarray image gridding with stochastic search based approaches. *Image and Vision Computing*, 25(2) :155 – 163, 2007.
- [GOOD-76] J. W. Goodman. Some fundamental properties of speckle. *J. Opt. Soc. Am.*, 66(11) :1145–1150, 1976.
- [HIRA-02] Roberto Hirata, Junior Barrera, Ronaldo F. Hashimoto, Daniel O. Dantas, and Gustavo H. Esteves. Segmentation of microarray images by mathematical. *Real-Time Imaging*, 8(6) :491–505, December 2002.
- [KATZ-02] M. Katzer, F. Kummert, and G. Sagerer. Robust automatic microarray image analysis. In *In BREW Bioinformatics Research and Education Workshop*, 2002.
- [KATZ-03] Mathias Katzer and Franz Kummert. A Markov random field model of microarray gridding. In *in Proc. ACM Symposium on Applied Computing (SAC)*, pages 72–77. ACM Press, 2003.

- [KEND-52] M. G. Kendall and A. Stuart. *The Advanced Theory Of Statistics*, volume 1. Charles Griffin and Compagny -Limited London and High Wycombe, third edition, 1952.
- [KULK-00] Alexander Kuklin, Soheil Shams, and Shishir Shah. Automation in microarray image analysis with autogene(tm). *Journal of the Association for Laboratory Automation*, 5(5) :67 – 70, 2000.
- [MCGO-02] M. McGoven and R. Fayek. Advantages of laser confocal microarray scanning. *Microarray Image Anlysis-Nuts and Bolts*, pages 51–68, 2002.
- [PAL-93] NR Pal and SK Pal. A review on image segmentation techniques. *Pattern Recognition*, 26 :1277–1294, 1993.
- [PET-02] A. Pretrov, S. Sha, S. Draghici, and S. Shams. Microarray image processing and quality control,. *Microarray Image Analysis-Nuts & Bolts*, (6) :99–130, 2002.
- [PET-04] A. Petrov and S. Shams. Microarray image processing and quality control. *Journal of VLSI Signal Processing Systems*, 38(3) :211–226, November 2004.
- [WEE-03] Alan Wee-Chung Liew, Hong Yan, and Mengsu Yang. Robust adaptive spot segmentation of dna microarray images. *Pattern Recognition*, 36(5) :1251 – 1254, 2003.
- [WIER-02] C. K. Wierling, M. Steinfath, T. Elge, S. Shculzen-Kremer, P. Aanstad, M. Clark, H. Lehrach, and R. Herwig. Simulation of dna hybridization experiments and evaluation of critical parameters during subsequent image and data analysis. *BMC bioinformatics*, 3(29) :17, October 2002.
- [YANG-02] Y. H. Yang, M. J. Buckley, S. Dudoit, and T. P. Speed. Comparison of methods for image analysis on cdna microarray data. *Journal of Computational and Graphical Statistics*, 11 :108–136, 2002.
- [YIDO-97] C. Yidong, R. Edward, and all. Ratio-based decisons and the quantitative analysis of cDNA microrarray images. *Journal of biomedical optics*, 2(4) :364–374, October 1997.